动态字典破解用户口令与口令选择 *

张学旺 [1,2]，孟 磊 [1]，周 印 [1]

(1. 重庆邮电大学 软件工程学院，重庆 400065; 2. 重庆大学 微电子与通信工程学院，重庆 400044)

摘 要：口令认证一直是最主要的身份认证方式。考虑到口令要满足口令策略和易记忆的要求，用户常常会将个人信息组合起来作为口令。为了调查此类口令的比例，以 2011 年泄露的 4 种真实口令集为实验素材，预先设定口令的组合结构和格式，使用程序统计使用个人信息组合作为口令的比例。实验结果表明，使用姓名、电话号码、特殊日期等信息组合而成的口令比例约为 12.41%~25.53%。根据这一规律，提出了动态字典攻击。攻击者可以在获得用户部分个人信息后，生成具有针对性的动态字词典，并以此来破解用户口令。最后，还讨论了如何选择口令以防止攻击者通过动态字典破解用户口令。

关键词：口令安全；动态字典攻击；口令选择；组合口令

中图分类号：TP309.2 doi: 10.19734/j.issn.1001-3695.2018.09.0752

Password cracking using dynamic dictionary and password selection

Zhang Xuewang[1, 2], Meng Lei[1], Zhou Yin[1],

(1. *School of Software Engineering*, *Chongqing University of Posts & Telecommunications*, *Chongqing* 400065, *China*; 2. *School of Microelectronics & Communication Engineering*, *Chongqing University*, *Chongqing* 400065, *China*)

Abstract: Password has always been the one of the most important way for identity authentication. to meet the requirements of password policy and memory, users often combine personal information as their passwords. Therefore, in order to investigate the proportion of such passwords. This paper used 4 real password sets leaked in 2011 as experimental materials, preset the combination structure and format of the password and used an application to calculate the proportion of the password combined by personal information. Experimental results showed that the proportion of passwords combined with names, phone numbers, special dates and other information was about 12.41% to 25.53%. According to this rule, dynamic dictionary attack is proposed. An attacker can generate a dynamic dictionary and use it to crack the user password, after he/she obtains some of the user's personal information. At last, this paper discusses that users how to select their passwords to prevent an attacker from cracking their passwords through a dynamic dictionary.

Key words: password security; dynamic dictionary attack; combinative password

0 引言

口令认证是最广泛使用的信息系统身份认证方法之一，尽管许多新型身份认证方法不断出现，口令面临着暴力破解和字典攻击等威胁，但口令具有易理解、易实现且成本低廉等优势，在可预见的未来仍将是最主要的身份认证方式[1]。

暴力破解和字典攻击是主流的口令破解方式。字典攻击逐一尝试自定义字典中的口令，直到尝试到正确的口令或者穷尽字典中的口令为止。字典攻击的成功率由字典中的口令决定，由于攻击者受限于时空矛盾，不能简单地增加字典中的口令数量来增加破解口令的成功率[2]。因此，研究人员试图找到一种方法来生成更精确的字典，从而在特定的时空消耗条件下提高破解口令的成功率。

对用户口令的攻击，根据攻击过程中是否利用用户个人信息，可将其分为漫步攻击和定向攻击[1]。漫步攻击(trawling attacking)是指攻击者不关心具体的攻击对象是谁，其唯一目标是在允许的猜测次数下，猜测出越多的口令越好。定向攻击的目标是尽可能以最快速度猜测出所给用户在给定服务（如网站，个人电脑）的口令。攻击者通常利用与攻击对象相关的个人信息，来增强猜测的针对性。因此，相对于漫步攻击，定向攻击更具有针对性，从而在在线攻击方面有明显的优势。

考虑到口令的可记忆性，用户往往将自身的信息通过某种构造方式形成口令，这样既方便了记忆，同时似乎让口令看起来复杂些（从格式上看，口令既有数字，也有字母等）。比如，长度为 12 的口令 Wang19991201，同时包含了数字、大写和小写字母；如果使用暴力破解方法，平均需要 $62^{12}/2$（只考虑数字，大小写字母的情况下）次才能破解。从这个角度看该口令几乎是绝对安全的。但是，如果了解了口令的构造方式（比如，“姓的拼音+特殊日期”），又获取了用户的个人信息，那么可能在经过极少的猜测后就能破解口令。暴力破解和不具有针对性的字典攻击，受猜测次数限制难以实施在线破解口令，因此具有针对性的字典攻击才具有实际应用意义。

本文提出一种定向的字典攻击方式，即动态字典攻击。在收集攻击对象的个人信息后，以某些常用格式组合用户信

收稿日期：2018-09-03；**修回日期：**2018-11-13 **基金项目：**国家自然科学基金资助项目（61571032）；重庆市重点产业共性关键技术创新专项重大主题专项项目（cstc2017zdcy-zdzxX0013）；重庆市教委人文社会科学科研重点项目（18SKGH033）

作者简介：张学旺（1974-），男，湖南祁东人，副教授，硕导，博士，主要研究方向为区块链与大数据、数据安全与隐私保护、网络通信软件等（zhangxw@cqupt.edu.cn）；孟磊（1993-），男，硕士研究生，主要研究方向为软件与网络安全、密码学；周印（1993-），男，硕士研究生，主要研究方向为跨媒体检索、大数据.

息生成动态字典，再用该字典攻击特定的用户口令。本文以真实的口令集为实验素材，分析了用户口令结构，展示了大约 12.41%~25.53%的口令是由用户的姓名、特殊日期、身份证号等个人信息组合而成的。还讨论了口令选择。用户会本能地选择用户友好型口令 [3]，但用户友好的口令通常非常脆弱；口令选择应在安全性和用户友好之间找到平衡点。

1 相关研究

由于普通字典攻击存在时空矛盾的问题，对于字典攻击的研究主要集中于使尽可能小的字典覆盖尽可能多的真实的口令[4]。近年来出现了基于真实的口令集的数据挖掘[5,6]，马尔可夫模型(Markov model) [7,8]，概率上下文无关模型(PCFG)和基于自然语言处理(natural language processing) [1,9]等方法。

2013 年，魏为民等人以 CSDN 和人人网泄露在网络上的口令集为实验素材[5]，做实验分析用户设置口令的习惯。获得了下列实验结果：用户喜欢用电话号码，特殊日期（比如生日、纪念日等）或者其他与用户名相关的信息作为口令；口令的长度主要集中在 8 到 9 位；全数字和数字+字母是最普遍的两种口令结构。还给出了一些在口令选择上的建议，但是这些建议大多没有考虑用户友好性，因此大多数用户不会采用这些建议。2016 年，刘功申等人[6]基于真实的口令集对口令进行了更加细致的统计与分析，同时比较了真实的口令集与黑客字典，发现即使是最著名的黑客字典对这些口令集的覆盖率都很低。可见没有针对性的字典攻击存在很大的提升空间。

基于马尔可夫模型的口令猜测，最早是由 Narayanan 等人[7]在 2005 年首次提出。它首先使用口令集进行训练，通过字符间的先后关系来计算口令的概率。把概率高的口令作为优先猜测的对象，从而在一定程度上解决了时空矛盾。2014 年，Ma 等人[8]首次将平滑技术和正规化技术运用到马尔可夫模型中，平滑技术主要是为了解决数据集中过度拟合问题。

基于概率上下文无关模型的口令猜测，由 Weir 等人在 2009 年提出[2]。和马尔可夫模型一样，基于概率上下文无关模型需要口令集进行训练。不同的是 PCFG 模型统计口令层面的结构化信息，诸如词性、句法等。对 PCFG 模型的改进主要集中在引入更多的规则使得口令生成的概率分布更加的合理。

2014 年，Veras 等人提出口令中包含了大量语义信息[9]，而这些信息是 PCFG 模型和马尔可夫模型难以利用到的；进一步地，在 PCFG 模型基础上提出了融合语义的 NLP 算法。

这些尝试提高破解概率的方法都是试图找到口令中的分布规律，“如果口令的分布确定了，那么破解口令的代价就会下降。” [10]但是，它们都没有融合个人信息，不存在针对性，难以实现在线的攻击。2016 年，Wang 等人[11]提出了基于马尔可夫链的定向猜测算法，首次将个人信息与马尔可夫结合。2016 年，Li 等人[12]首次提出了基于 PCFG 的定向攻击猜测算法，即利用攻击对象相关的个人信息。但是 Wang，Li 对个人信息的组合分析不够全面，比如，作者仅提到了姓+生日的组合，对姓名+生日、姓名拼音首写+生日、姓+名的拼音首写+生日等格式没有考虑。

2 用户口令分析

本文在互联网上获取了 2011 年泄露的真实口令集，这些口令集主要是 CSDN、人人网、嘟嘟网和多玩网的用户账号和口令信息。虽然这些网站的信息有数百万条，但是其中一些口令长度太短，不符合如今的现实情况，因此实验排除了长度小于 6 的口令。除此之外，还有些无效的口令来源于“网络水军”。“网络水军”通常在同一时间段注册，它们常常会有相同的账号或者口令。根据这些特点，本文使用程序清理这些无效的用户信息。口令集的信息如表 1 所示。

表 1 口令集信息

Table 1 Information of password sets

网络名	总数量	有效数量	网站类型
CSDN	6421280	6375660	IT 社区
人人网	4421776	4178592	社交媒体
嘟嘟网	14108052	9905651	网页游戏
多玩网	2635064	2583388	游戏

2.1 用户信息类型与格式

用户可以组合个人信息作为口令，口令可以由姓名、特殊日期、电话号码、社交媒体账号、爱好、偶像名等组成。因为信息收集的难度问题，本实验仅关注姓名、特殊日期、电话号码、社交媒体账号、身份证最后 6 位和长度小于等于 4 的特殊数字。日期和名字有多种格式，例如，用户的生日是 1999 年 6 月 8 日，它可以写成 19990608，990608 等。同样，用户姓名也有多种格式，表 2 和 3 分别是日期和姓名主要的格式。

表 2 日期格式

Table 2 Format of date

	日期：1999 年 6 月 8 日
格式 1	19990608
格式 2	990608
格式 3	199968
格式 4	9968
格式 5	1999
格式 6	0608

表 3 姓名格式

Table 3 Format of name

	姓名：林智强
格式 1	linzhiqiang, LinZhiQiang, Linzhiqiang
格式 2	linzhiq, LinZhiQ, Linzhiq
格式 3	linzq, LinZQ, Linzq
格式 4	zqlin, ZQLin, zqLin
格式 5	lzq, LZQ, Lzq
格式 6	lin, Lin

这些合适可以作为判断口令的一部分是否为姓名或者日期的标准。因为汉语拼音的数量有限，如果口令的一部分由拼音组成，再考虑到它的可记忆性，那么基本肯定它有意义；或者是名字，或者是特定汉语词汇。同样，如果一串 11 位数字是以 130、133 等开头，那么可以基本确定它是电话号码。130、133 等是网络识别号，如表 4 所示，表中给出了中国目前所有的网络识别号。对于社交媒体，实验主要以 QQ 为例，但是，单从号码上很难直接判断一串 6 到 11 位的数字是否为 QQ 号码。同样，根据中国居民身份证的编号特征，也很验证判断一串 6 位的数字是否为身份证后 6 位数字。所以，本次实验中基本不能断定一串数字是否为 QQ 号码，或者是身份证后 6 位数字。

表 4 网络识别号

Table 4 Network identification number

电信运营商	网络识别号
中国移动	134、135、136、137、138、139、147、150、151、152、157、158、159、178、182、183、184、187、188、198、1705
中国电信	133、153、177、180、181、189、199，1700
中国联通	130、131、132、155、156、185、186、145、176、166，1709

2.2 口令结构

用户将个人信息组合起来作为口令，有一些组成结构。本文给出一些比较常见的结构以研究它们的频率。如表 5 所示，以字母 A 表示口令中的字母串或字母，字母 D 表示口令中的数字字符串，字母 N 表示口令中的拼音部分（大部分为姓名的拼音格式）。其中 A 主要包括如 qq、abc、aaaa 等，D 主要包括电话号码、QQ 号码、特殊日期、身份证号后 6 位数字、长度小于等于 4 的数字字符串。在实验中，编写程序完成判断口令结构的工作。

表 5 口令结构

Table 5 Password structure

	口令结构
格式 1	A + D
格式 2	A + D + A
格式 3	N + D

3 实验结果与分析

3.1 实验结果

本文已经列出了口令的结构和口令格式的判断标准。根据上述口令分析，本文编写程序以分析判断出哪些口令是由用户的个人信息组合而成的口令。实验统计结果如表 6 所示。其中 L 表示口令中的字母，N 表示口令中的姓名、汉语词汇的拼音，SD 表示口令中的特殊日期，PN 表示口令中的电话号码，QQ 表示口令中的 QQ 号码，LS4 表示口令中长度小于等于 4 的数字字符串，ID6 表示口令中身份证号后 6 位数字。

表 6 实验统计结果

Table 6 Experimental results

口令结构	CSDN		人人网		嘟嘟网		多玩网	
	数量	比例	数量	比例	数量	比例	数量	比例
L+PN	5116	0.08%	167	0.04%	3445	0.03%	15187	0.59%
L+SD	11887	0.19%	22467	0.54%	277880	2.81%	11020	0.43%
L+ID6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
L+QQ	122053	1.91%	40536	0.97%	873764	8.82%	296486	11.48%
L+PN+L	364	0.01%	7	0.00%	69	0.00%	399	0.02%
L+SD+L	2625	0.04%	936	0.02%	4679	0.05%	1448	0.06%
L+ID6+L	24883	0.39%	17135	0.41%	24879	0.25%	12680	0.49%
L+QQ+L	78824	1.24%	17990	0.43%	71373	0.72%	72061	2.79%
N+PN	22845	0.36%	442	0.01%	3171	0.03%	16124	0.62%
N+SD	431088	6.76%	141909	3.40%	639413	6.50%	214465	8.30%
N+LS4	317206	%4.98	359821	8.61%	1039157	10.49%	400803	15.51%
N+ID6	252575	3.96%	44204	1.01%	241088	2.43%	54300	2.10%
N+QQ	590652	9.26%	91546	2.20%	961787	9.71%	459434	17.78%
PN	189340	2.97%	159517	3.82%	382070	3.86%	17	0.00%
SD	409284	6.42%	315328	7.55%	392482	3.96%	6	0.00%
ID6	28134	0.44%	366982	8.78%	435835	4.40%	3	0.00%
QQ	1263310	19.81%	1241562	29.71%	2623371	26.48%	406	0.02%
N	276633	4.33%	344951	8.26%	408231	4.12%	154801	5.99%

由于 QQ 号码和身份证后 6 位数判断不准确，不计入统计分析，本文主要分析结构为 L+PN、L+SD、L+PN+L、L+SD+L、N+PN、N+SD 和 N+LS4 的口令；把 PN、SD、N 仅含单一用户信息的口令定义为单信息口令，本文的程序统计的有效的用户个人信息组合口令比例如表 7 所示。实验结果表明，有效的用户个人信息组合口令的比例从 12.41%到 25.53%。

表 7 有效个人信息组合口令比例

Table 7 Proportion of password combined with valid personal information

	CSDN	人人网	嘟嘟网	多玩网
比例	12.4087%	12.5820%	19.8656%	25.5264%

3.2 实验结果分析

口令强度与用户的性别、专业、年龄都有着某种联系。Mazurek 等[13]的实验结果表明，相对其他专业，计算机科学专业的学生设置的口令强度是最强的。CSDN 全称是中国软件开发者网站，CSDN 的大部分用户是计算机专业或者计算机相关专业人员，CSDN 的有效的个人用户信息组合口令的比例在实验口令集中最低。多玩网和嘟嘟网都是游戏网站，但多玩网的有效的个人用户信息组合口令的比例在实验口令集中最高，要明显高于嘟嘟网；导致这个结果的原因与网站的口令策略有关。多玩网的单信息口令中除结构为 N 的口令比例不为 0 以外，其余的几乎都为 0。因此本文可以判断多玩网禁止用户设置全为数字的口令。如图 1 所示，当单一信息口令的比例下降时，有效个人用户信息组合口令的比例会增加。除了 CSDN 以外，其余的三个网站的有效信息组合口令与单一信息口令之和基本都为 30%，也就是约 30%的口令都是由用户个人信息组合构成。

用户可以用姓名、生日和其他信息来构建个人的口令[3,14,15]。现在，许多网站或应用程序都有口令策略，以防止用户设置弱口令。用户可以将一些信息作为口令组合，以满足口令策略，同时又方便用户记忆。从表面上来看，这些组合口令是要比单信息口令的安全性更高些，但当这个规律被发现后，安全性也是比较脆弱的。攻击者可以根据这个规则破解用户的口令。攻击者收集关于目标用户的一些信息，这些信息包括他/她的姓名、电子邮件地址、ID 号等等；使用社会工程学来收集关于用户的信息[16]。例如，可以通过被丢弃的快递的包装盒来获得目标用户的电话号码和淘宝账号（账户也极有可能是他的电子邮箱号）。攻击者利用收集到的信息制作动态字典。使用 John The Ripper 等软件和动态字典破解目标用户的口令。该方法针对性强，效率高。

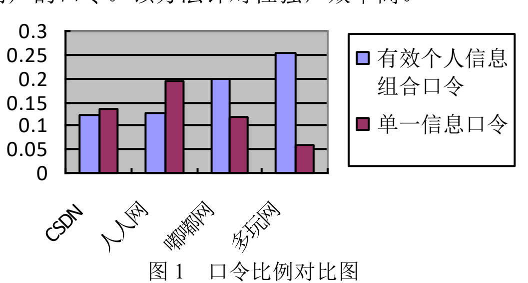


图 1 口令比例对比图

Fig. 1 Password proportion comparison chart

根据实验结果，接近 12.41%到 25.53%的口令由用户个人信息组合而成的。因此，搜集了攻击对象的信息，就可以快速地按照表 6 中的口令结构生成组合口令、再生成动态字典。如表 8 所示，显示了有效个人信息组合口令的 7 种口令结构的最大猜测次数。可以发现，动态字典中的口令条数在 10^5 的量级就可以覆盖所用结构的信息组合口令，也即最多猜测约 $1.7*10^5$ 次就有 12.41%到 25.53%的概率猜测出用户的口令。本文的实验还未找到合适的方法来精确地判断 QQ 号码和身份证号后 6 位，所以不知道口令结构为 L+QQ、L+QQ+L、N+QQ、L+ID6、L+ID6+L 和 N+ID6 的比例。但是，可以肯定的是，有些口令的确是由 QQ 号码、身份证号后 6 位数字和其他信息组成的。因此，如果比较全面的搜集目标的用户的个人信息来构建动态字典，并以此来猜测用户

口令的概率要高于 12.41%到 25.53%。

表 8 有效个人信息组合口令的 7 种口令结构最大猜测次数

Table 8 The maximum number of guesses of seven password structures for valid personal information combination passwords

口令结构	最大猜测次数
L+PN	26
L+SD	26*6
L+PN+L	26*26
L+SD+L	26*26*6
N+PN	17
N+SD	17*6
N+LS4	$17*10^4$

4 口令选择

网站或应用程序在用户设置口令时设定一些限制，可以增强口令的强度。根据本文的实验结果，用户可以组合一些个人信息来设置口令，但是攻击者也可以使用动态字典破解口令。那么如何选择口令以避免使用动态字典破解用户口令呢？

用户使用单一信息作为口令是很容易被猜测到的。Alvare 提出了一些设置口令的方法原则[4]。这些方法教用户如何选择他们的口令，所得到口令看起来很强大，但事实并非如此。教用户使用某种方法或者原则设置口令本身是危险的。例如，如果为用户提供了创建口令的特定方式，比如使用最喜欢的短语或者名言的首字母，如果许多用户都使用了这种方法来生成口令，那么这些口令将会形成规律性，这样也就能通过特定的程序破解。

大多数网站和应用程序都使用了口令策略，以防止用户设置弱口令。所以，似乎更严格的口令策略，就会有更强的口令，显然严格的口令策略会产生复杂的口令，复杂性增加了口令对攻击的抵抗力，但降低了可用性[17]。用户很难记住他们的口令，所以，他们可能会把口令写下来[3,13]，从而导致新的危险。与此同时，严格的口令策略会让用户抱怨，甚至会放弃注册[3]。对于一个网站或应用程序来说，吸引更多的固定用户是很重要的，严格的口令策略可能会导致更大的损失。

安全性和可记忆性对口令都很重要。王秀利对国内用户口令的安全性和可记忆性进行了定量分析[18]，提出了能根据用户历史口令数据动态约束用户口令设置行为的口令创建规则：若用户采用纯数字口令，则口令长度不应低于 7 位；大写字母+小写字母组合的口令应避开 6 位和 8 位；大写字母+特殊字符的组合推荐使用 9 位。这个规则对口令选择有帮助，但它没有考虑按照这些规则创建的口令的含义。有意义的口令更容易记住[19]。

当用户选择口令时，应该考虑口令的安全性和可记忆性。用户通过组合个人信息来创建口令是很正常的，因为它很容易记住。关键是用户应该使用什么样的信息。一些信息，如用户的姓名、生日、身份证号码和社保号码，这些信息很容易获得。用户应该使用秘密信息，比如上小学时的教室门牌号、你的第一台电脑或汽车的价格等等。用户可能有“幸运数字”，用户可以使用“幸运数字”作为口令的一部分。总之，用作创建口令的信息应该是保密的，不可轻易获取的。最后，本文提出一些口令创建与使用的建议。

a) 口令的长度不应该太短，应至少为 8。

b) 口令不应该写下来或存储在你的电脑里。

c) 不要在没有加密的网络环境中传输口令。

d) 尽量为每个应用账号设置不同的口令。

e) 不要使用网站或应用程序的“记住口令”功能。

5 结束语

本文借助互联网获取了 2011 年泄露的真实口令集（CSDN、人人网、嘟嘟网和多玩网），并编写程序删除了口令集中无效的口令。以这些口令集为基础，本文通过实验方式提出了动态字典攻击方法。在实验中，归纳总结了主要的口令结构，分析有效的个人信息组合口令的 7 种口令结构并给出了这 7 种口令结构的有效用户个人信息组合口令的比例。攻击者可以利用这个规律，首先获取目标用户的个人信息，然后根据这些信息生成动态字典，再通过动态字典破解用户的口令。

用户喜欢将个人信息组合成口令，同时口令的可记忆性对用户也很重要。本文基于口令的安全性和可记忆性，对口令选择给出了一些建议，建议用户组合不易获取的秘密个人信息以创建口令。

参考文献：

[1] 王平，汪定，黄欣沂．口令安全研究进展 [J]．计算机研究与发展，2016, 53 (10): 2173-2188. (Wang Ping, Wang Ding, Huang Xinyi. Advances in password security [J]. Journal of Computer Research and Development, 2016, 53 (10): 2173-2188.)

[2] Weir M, Aggarwal S, Medeiros B D, *et al.* Password cracking using probabilistic context-free grammars [C]// Proc of the 30th IEEE Symposium on Security and Privacy. Piscataway,NJ: IEEE Press, 2009: 391-405.

[3] Alvare A M D. How crackers crack passwords or what password to avoid [C]// Proc of the 2nd USENIX Security Workshop, Berkeley:USENIX, 1988: 103-112.

[4] 周浩，王靖康，王博，等．明文口令生成模型研究综述 [J]．计算机工程与应用, 2018,54(4): 9-16. (Zhou Hao, Wang Jingkang, Wang Bo, *et al.* Comprehensive overview of plaintext password generation models [J]. Computer Engineering and Applications, 2018 ,54(4): 9-16.)

[5] 魏为民，陈为召，李红娇．国内网络用户密码分析 [J]．上海电力学院学报, 2013, 29 (6): 584-588, 606. (Wei Weimin, Chen Weizhao, Li Hongjiao. Analyzing network password habits [J]. Journal of Shanghai University of Electric Power, 2013, 29 (6): 584-588, 606.)

[6] 刘功申，邱卫东，孟魁，等．基于真实数据挖掘的口令脆弱性评估及恢复 [J]．计算机学报, 2016, 39 (3): 454-467. (Liu Gongshen, Qiu Weidong, Meng Kui, *et al.* Password vulnerability assessment and recovery based on rules mined from large-scale real data [J]. Chinese Journal of Computers, 2016, 39 (3): 454-467.)

[7] Narayanan A, Shmatikov V. Fast dictionary attacks on passwords using time-space tradeoff [C]//Proc of ACM Conference on Computer and Communications Security. New York: ACM Press, 2005: 364-372.

[8] Ma J, Yang W, Luo M, *et al.* A study of probabilistic password models [C]// Proc of IEEE Symposium on Security and Privacy.Piscataway,NJ: IEEE Press, 2014: 689-704.

[9] Veras R, Collins C, Thorpe J. On semantic patterns of passwords and their security impact [C]// Proc of Network and Distributed System Security Symposium. San Diego: Internet Society Press, 2014: 1-16

[10] Malone D, Maher K. Investigating the distribution of password choices [J]. Computer Science, 2012, 8 (3): 301-310.

[11] Wang D, Zhang Z, Wang P, *et al.* Targeted online password guessing: an underestimated threat [C]//Proc of ACM SIGSAC Conference on

Computer and Communications Security. New York: ACM Press, 2016: 1242-1254.

[12] Li Y, Wang H, Sun K. A study of personal information in human-chosen passwords and its security implications [C]// Proc of the 35th Annual IEEE International Conference on Computer Communications. Piscataway,NJ: IEEE Press, 2016: 1-9

[13] Mazurek M L, Komanduri S, Vidas T, *et al.* Measuring password guessability for an entire university [C]// Proc of ACM SIGSAC Conference on Computer & Communications Security. New York: ACM Press, 2013: 173-186.

[14] Klein D V. A survey of, and improvements to, password security [J]. Programming & Computer Software, 2001, 17 (3): 5-14.

[15] Spafford E H. Observing reusable password choices [C]// Proc of the 3rd Security Symposium. Berkeley:Usenix Press, 1992: 299-312.

[16] Diane B M S S, Faith M M S. Password security: an empirical investigation into E-commerce passwords and their crack times [J]. Information Systems Security, 2006, 15 (6): 45-55.

[17] Herley C. Where do security policies come from? [C]// Proc of the 6th Symposium on Usable Privacy and Security. New York: ACM Press, 2010: 10.

[18] 王秀利. 基于真实数据集的密码定量分析及规则创建 [J]. 信息网络安全, 2015 (12): 42-47. (Wang Xiuli. Quantitative analysis and create policy of password based on real dataset [J]. Netinfo Security, 2015 (12): 42-47.)

[19] Woods N, Siponen M. Too many passwords?How understanding our memory can increase password memorability [J]. International Journal of Human-Computer Studies, 2018 (111): 36-48.